Proc. Assoc. Advmt. Anim. Breed. Genet. 22:121-124

QUANTITATIVE GENOMIC ANALYSES IN THE PACIFIC WHITELEG SHRIMP *LITOPENAEUS VANNAMEI*

M.S Khatkar[1,2], K.R. Zenger[2,3], D.B. Jones[2,3], Jeff Prochaska[4], H.A.M van der Steen[4], D.R. Jerry[2,3], and H.W. Raadsma[1,2]

[1] Sydney School of Veterinary Science, Faculty of Science, The University of Sydney, Camden, NSW, Australia
[2]ARC Hub for Advanced Prawn Breeding, James Cook University, Queensland, Australia
[3] Centre for Sustainable Tropical Fisheries and Aquaculture, College of Science and Engineering, James Cook University, Queensland, Australia
[4] Global Gen, Jalan Raya Narogong Km 14, Desa Cikiwul Bantar Gebang Bekasi, 17310 Indonesia

SUMMARY

Traditional genetic improvement programs for Pacific white-leg shrimp (*Litopenaeus vannamei)* rely on family selection to improve growth and disease resistance traits. DNA technologies can help in simplifying breeding schemes and increasing genetic gains particularly for complex or difficult to measure traits. Here we present the results of genome-wide association and whole genomic prediction analyses using average family allele frequencies and the family mean of a growth trait in a genetic resource population consisting 1,934 animals and 690 families of *L. vannamei* genotyped with 8,967 genome-wide SNPs. After correcting for FDR, no significant SNPs were detected for growth. The accuracy of DGV in mirror prediction is much higher (0.65-0.69) as compared to forward prediction. A SNP that may be closely linked to the sex locus was identified with the female being the heterogametic sex.

INTRODUCTION

The Pacific white-leg shrimp (*Litopenaeus vannamei)* is an important aquaculture species and the most widely farmed shrimp globally. Traditional genetic improvement programs for *L. vannamei* rely on family selection to improve growth and disease resistance traits. Recent advances in high-throughput genotyping and analytical methods can help to simplify breeding schemes and increase genetic gain, particularly for complex or difficult to measure traits. In particular the mapping of quantitative trait loci (QTL), or genes with large effect may have an immediate application in marker assisted selection (MAS). We conducted a genome-wide association analysis for growth and a sex associated trait in *L. vannamei* by genotyping a resource population with a purpose built genome wide SNP panel and explored the possibility of genomic selection in *L. vannamei*.

MATERIALS AND METHODS

We built a resource database for *L. vannamei* by genotyping a total of 1,934 samples with 8,967 genome-wide SNPs on the Illumina Infinium ShrimpLD-24 v1.0 genotyping array (Jones et al. 2017 - these proceedings). These included 1,134 female and 123 male parents along with 677 nauplii (larval shrimp) pools. Following SNP quality control (QC), 5,893 SNPs were used for all analyses. An integrated linkage and LODE map was constructed using 631 progeny from 30 grand maternal and 19 grand paternal traced families (Jones et al 2017 - these proceedings). In total, 4,817 SNPs were mapped to 44 linkage groups that span a total of 4552.5 cM and cover an estimated 98.12% of the *L. vannamei* genome. The average interval, excluding intervals of 0 cM, was 2.67 cM. This map was utilised for all subsequent GWAS analyses and presenting results as Manhattan plots.

For the GWAS, average family allele frequencies were used for 690 families. For an additional 94 families, the genotype of the parents were available and for these the realised family-mean allele frequencies were computed as the mean of parental alleles. Out of these, based on the availability of

genotypic and phenotypic data, family-wise mean frequencies on 416 families were finally used for conducting SNP association with the traits.

Phenotypic data on the family mean of 416 families on one growth trait G.d2All (Growth rate in grams per day for all tanks) were used for the current analyses. The family mean value of G.d2All ranged from 0.16 to 0.45 with a mean and standard deviation of 0.31 and 0.05, respectively. The overall distribution of G.d2All (Figure 1a) indicated that this trait is normally distributed. Figure 1b shows batch-wise distribution. In addition, individual genotypes of 1,963 animals and their sex-status were analysed to detect any sex associated SNP association.

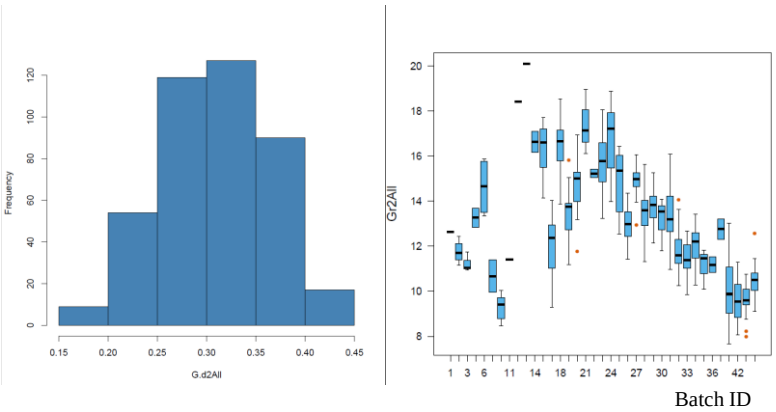

Figure 1. a) Overall distribution of the growth trait, G.d2All, presented as histogram. b) Batch-wise mean and distribution of the growth trait. The x axis represents batch id in a chronological order.

Genome-wide association (GWA) analyses. The association analysis was conducted using the allele frequencies and mean phenotypic value of the traits for the families. A realized additive relationship matrix (**K**) (Endelman, 2011) was computed to calculate molecular kinship among all families using scaled mean allelic frequencies. The regression of the mean family phenotype on SNP genotypes were conducted by fitting the mean allele frequency as a covariate and adjusting for across family relationships using the following linear mixed model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon}$$

where **y** is a vector of the phenotypic value (trait), **X** is the incidence matrix incorporating mean and SNP allele frequency; **β** is a vector representing coefficients of the fixed effects, **Z** is an incidence matrix mapping phenotype records to families, **u** is a vector of polygenic genetic effects such that $\mathrm{var}(\mathbf{u}) = s_g^2\mathbf{K}$, where **K** is the kinship matrix as described above, and **ε** is vector of residual random errors with $\mathrm{var}(\boldsymbol{\varepsilon}) = s_e^2\mathbf{I}$. The model was fitted using ASReml (Gilmour, Gogele, Cullis, & Thompson, 2009). Genome-wide false discovery rate was computed using the q-value package in R (www.r-project.org).

Accuracy of genomic prediction. Genomic selection uses information from all SNP to derive Direct Genomic Values (DGV). Accuracy of direct genomic values using SNP genotypes was investigated by dividing the data on 416 families into a training set and a validation/test set. Three different sets of training and test sets were investigated by using a different proportion of the families in validation and test sets viz. 1) 75 % in training and 25 % in test; 2) 67% in training and 33 % in test; 3) 50 % in training and 50 % in test. In forward prediction, the training set consisted of older

families and the test set consisted of recent families. In the mirror prediction the families were allotted randomly to the training and the test sets across all batches.
DGV were estimated using a best linear unbiased prediction (BLUP) method which used a Gaussian kernel prediction based on the Euclidean distance matrix for K where K is the kinship matrix as described above. This model is implemented in R package rrBLUP (Endelman, 2011). The accuracy of DGV prediction was computed as the Pearson's correlation coefficient between DGV and the mean phenotypic value of the families in the test set. The bias was computed as the regression coefficient of DGV on the phenotypic value.

RESULTS AND DISCUSSION

Genome-wide association (GWA) analyses for growth. Genome-wide associations expressed as log-P value for each marker are presented as a Manhattan plot in Figure 2. The unmapped SNPs are shown without any chromosome label on right hand side of the plot. There were 83 SNPs significant at P <0.05. However, after correcting for FDR no significant SNPs were detected for growth. A few clusters of SNPs with P-value <0.001 were identified, however, due to the high FDR, these could only be considered as suggestive at best. Overall these GWAS results suggest that no gene of large effect regulates this growth trait. In order to detect significant SNPs of moderate or small effect, a substantially larger sample size and a higher SNP density would be required.

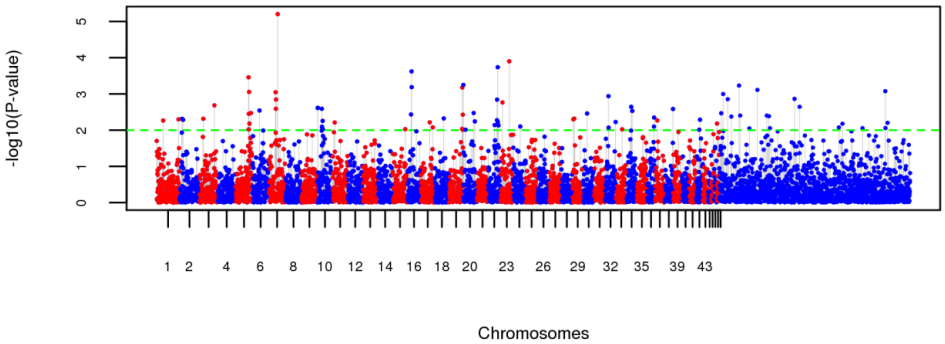

Figure 2: Genome wide SNPs associations with growth trait, G.d2All, presented as Manhattan plot.

Sex-associated SNPs. The genome-wide associations of SNPs with sex status of the animals presented in Figure 3 as Manhattan plot shows one very significant cluster of SNPs on linkage group 44. The most significant SNP was associated with sex status of the animals with -log10 (p)=294 at the start of LG44. Minor allelic frequency for this SNP was 0.3 indicating that this is a common SNP. The strong association and frequency of males and females genotypes suggest that this SNP may be closely linked to the sex locus. Most females (95%) were heterozygous whereas most males (95%) were homozygous for the major allele of the most significant SNP. These results are in agreement with earlier studies which suggest that the sex in penaeid species is mainly genetic and determined by a WZ–ZZ chromosomal system where the female is the heterogametic sex (Staelens et al., 2008). This also raised possibility of monosex culture. Potentially homogametic females and males can be used as parents to yield sexually uniform heterogametic female offspring. Monosex sex culture in prawn has been reported more profitable as compared to rearing of mixed sex animals (Mohanakumaran Nair, Salin, Raju, & Sebastian, 2006). In addition, monsex culture system can provide some protection to genetically superior stock.

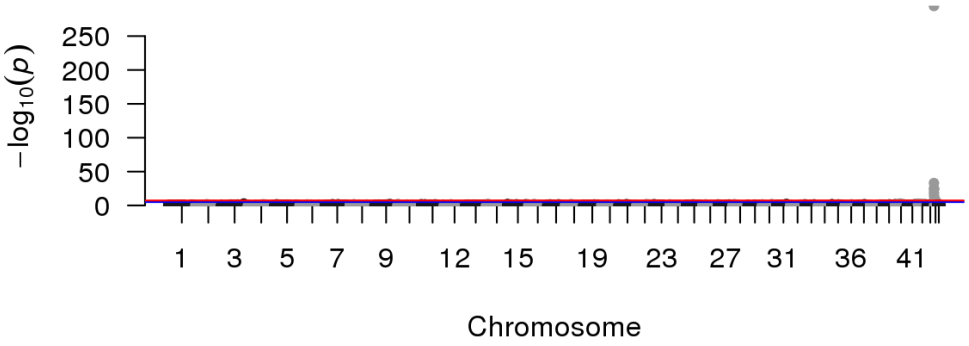

Figure 3. Genome-wide associations for sex status of animals.

Accuracy of genomic prediction of growth trait (G.d2All). The data on a growth trait G.d2All on 416 families were analysed for this analysis. The mean of each family was obtained by pooling data across tanks as described in the methods section. The genotypic data on family-wise mean allelic frequencies for 4,686 QC SNPs were included in this analysis.

The accuracy of DGV in mirror prediction (randomly dividing families in training and test set, Table 1) is much higher (0.65-0.69) as compared to forward prediction (0.17-0.32) (Table 1). The prediction accuracies in the mirror prediction indicate the potential level of accuracies of genomic selection in shrimp. It seems that declining trend with a very large batch effect of G.d2All (Figure 1b) hampered the accuracy of genomic prediction in forward prediction.

Partitioning 50 % families in training and 50 % in test gave higher accuracy as compared to other partitions in forward prediction (Table 1). Inconsistent accuracies in different partition/sets indicate that the current sample size for genomic prediction is too small which is further complicated by the large batch effect confounded with families.

Table 1. Accuracy of genomic prediction for a growth trait (G.d2All).

% in training	Number of families		Mirror prediction		Forward prediction	
	Training	Test	Accuracy	Bias	Accuracy	Bias
75	312	104	0.693	1.256	0.168	0.413
67	277	139	0.632	1.039	0.279	0.757
50	208	208	0.647	1.478	0.315	0.671

CONCLUSION

This study identified a major region associated with sex, and demonstrated that genomic selection has potential application with moderate number of SNPs, family average phenotypic records, and based on family DNA pool frequency data for commercially important traits in *L. vannamei*.

REFERENCES

Endelman, J. B. (2011). The Plant Genome, **4**(3), 250-255.

Gilmour, A. R., Gogele, B. J., Cullis, B. R., & Thompson, R. (2009). ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Mohanakumaran Nair, C., Salin, K. R., Raju, M. S., & Sebastian, M. (2006). Aquaculture Research, **37**(9), 949-954.

Staelens, J., Rombaut, D., Vercauteren, et al. (2008). Genetics, **179**(2), 917-925.